**AUTOMATED MINING OF CROSS-CUTTING CONCERNS FROM PROBLEM REPORTS AND REQUIREMENTS SPECIFICATIONS**

Horatiu Dumitru, Adam Czauderna, and Jane Cleland-Huang*, School of Computing, DePaul University, Chicago, IL, 60604, jhuang@cs.depaul.edu

Detecting cross-cutting concerns early in the software development life-cycle offers significant benefits during the process of requirements analysis and architectural design. Instead of focusing purely on functional requirements, the architect is able to design a system that more fully satisfies stakeholders' needs through early identification of cross-cutting concerns. Without the benefit of early detection, cross-cutting concerns tend to go unnoticed and become interwoven throughout the design, inhibiting developers from identifying an optimal architectural design, and resulting in the need for later refactoring efforts. Several approaches have been suggested for mining cross-cutting concerns, also known as early aspects, from requirements; however prior techniques have been predominantly labor intensive or have required manually created training sets of data. In this work we introduce a new approach for identifying early aspects based on iterative clustering. During each iteration a consensus clustering algorithm is used to partition the requirements into meaningful and cohesive topics. Cohesion and size metrics are computed for each of the generated clusters and used to estimate the quality of each cluster. These values are combined in a quality metric and the cluster exhibiting the highest quality score is selected. Its core terms are extracted by considering the distance of all terms to the centroid of the cluster, and selecting the closest ones. These terms represent a newly identified candidate aspect and are placed into a list. In preparation for the next iteration, the core terms from the identified aspect are systematically removed from all requirements in the dataset so that future clusterings will form around different themes. The process is repeated iteratively until a stopping condition is reached which suggests that no additional quality topics can be identified. The approach was evaluated against over 1000 requirements for an airport kiosk system, and against the problem reports generated through the maintenance of a military airplane.

**Acknowledgment**

This work was partially funded by REU supplement grant # 0936416.